Disconnected contributions to hadronic structure: a new method for stochastic noise reduction

Sara Collins
Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany
E-mail: sara.collins@physik.uni-regensburg.de

Gunnar Bali
Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany
E-mail: gunnar.bali@physik.uni-regensburg.de

Andreas Schäfer
Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany
E-mail: andreas.schaefer@physik.uni-regensburg.de

We present a new method for reducing the stochastic noise of all-to-all propagators based on stopping the inversion of the propagator before convergence. The method is easy to implement, unbiased and independent of the quark action. Applying this method to the calculation of disconnected loops needed for hadronic structure observables we find savings in computer time of factors of 4 – 12 depending on the operator inserted in the loop. When combined with a hopping parameter expansion technique we obtain combined gains of up to factors of 30 for some operators.

The XXV International Symposium on Lattice Field Theory
July 30 - August 4 2007
Regensburg, Germany

*Speaker.
1. Introduction

Nucleon structure observables such as baryon form factors and moments of (generalised) parton distributions are extracted from 3pt functions which have connected and disconnected contributions. The latter, of the form $\text{Tr}(\Gamma M^{-1}) \times 2$pt function, are normally omitted as they require the calculation of all-to-all propagators $M_{i,j}^{-1}$. Instead often differences between observables, for example $g_A = \Delta u - \Delta d$, are quoted. However, any settling of the question of the spin or the strangeness content of the nucleon requires a calculation of the corresponding disconnected loops. In the following we present a new method for calculating all-to-all propagators which reduces the associated computational overhead of performing the subtraction was not significant.

1.1 Stochastic methods for all-to-all propagators

The standard method for computing all-to-all propagators is via stochastic sampling. A set of random complex $Z(2)$ noise vectors, $|\eta^l\rangle$, $l = 1 \ldots L$, is generated for which,

$$
\frac{1}{L} \sum_l |\eta^l\rangle\langle\eta^l| = 1 + O(1/\sqrt{L}), \quad \frac{1}{L} \sum_l |\eta^l\rangle = O(1/\sqrt{L}). \tag{1.1}
$$

Using these vectors as sources one can construct an unbiased estimate of the all-to-all propagator, $E_L(M^{-1})$, using $|\eta^l\rangle$ and the corresponding solution vectors $|s^l\rangle = M^{-1}|\eta^l\rangle$:

$$
E_L(M^{-1}) = \frac{1}{L} \sum_l |s^l\rangle\langle s^l| = M^{-1} + M^{-1} \left( \frac{1}{L} \sum_l |\eta^l\rangle\langle\eta^l| - 1 \right). \tag{1.2}
$$

From eqns. 1.1 and 1.2 it is clear that the stochastic error on the estimate only depends on the off-diagonal elements of $\frac{1}{L} \sum_l |\eta^l\rangle\langle\eta^l|$ and falls off as $O(1/\sqrt{L})$. For a fixed number of configurations, depending on the quantity studied (obtained using $E_L(M^{-1})$), the stochastic noise can dominate over the gauge noise and additional noise reduction techniques are required.

It is important to note that any noise reduction techniques should be unbiased and the resulting reduction in noise should justify the computational overhead. Existing techniques include “partitioning” [1] where each noise vector $|\eta^l\rangle$ is replaced by a set of partitioned vectors $|\eta^l_p\rangle$, $p = 1 \ldots P$, where $|\eta^l_p\rangle$ has many zeros. By zero-ing entries in the source vector one will avoid some of the large off-diagonal elements contributing in eqn. 1.2, the hope is that a smaller variance is obtained for the same amount of computer time despite $P$ times as many inversions. Wilcox [1] found the gain (in terms of computer time) for $\text{Tr}(\Gamma M^{-1})$ for colour-spin partitioning to depend strongly on $\Gamma$ but could be in the region of factors of $3 - 7$ (for $\Gamma = \gamma_{\mu\nu}$ and $\gamma_5 \gamma_\lambda$) or higher (for $\gamma_\lambda$).

The Kentucky group [2] take a different approach and use the hopping parameter expansion (HPE), to construct traceless estimates of the off-diagonal elements in eqn. 1.2. Subtracting these estimates from $\text{Tr}(\Gamma M^{-1})$, where $M = 1 - \kappa D$, leaves the trace unchanged but reduces the variance. This approach should work well in the heavy quark regime, for example for masses down to the strange quark mass. Mathur and Dong [3] found a gain of a factor of $6 - 7$ subtracting up to $\kappa^4 D^4$ for the strangeness contribution to the magnetic moment of the nucleon $G_{\mu\nu}(0)$. The computational overhead of performing the subtraction was not significant.
Additional approaches also exist: for example in the light quark regime one can calculate the low lying eigenmodes of the Dirac operator and use these to estimate part of the propagator \[^4\]. The remainder can be calculated stochastically \[^5\]. Different methods can often be combined.

### 1.2 A new approach: unbiased truncation of the solver

We present a new method for noise reduction which involves stopping the inversion of the stochastic propagator before convergence, i.e. using \(n_t\) iterations in the solver to obtain \(|s^r_{n_t}\rangle = M_{n_t}^{-1} |\eta^l\rangle\) instead of running to convergence using \(n_c\) iterations and obtaining \(|s^r_{n_c}\rangle = M_{n_c}^{-1} |\eta^l\rangle\). The difference between \(M_{n_t}^{-1}\) and \(M_{n_c}^{-1}\) can be estimated stochastically using an independent set of sources:

\[
E[M_{n_t}^{-1}] = E_{L_1}[M_{n_t}^{-1}] + E_{L_2}[M_{n_t}^{-1} - M_{n_c}^{-1}].
\]

(1.3)

This is based on an exact linear decomposition and the algorithm with which both parts are calculated is well defined. Using two independent sets of noise vectors for the two parts then implies that the stochastic error can be reduced by performing a large number of cheap inversions for \(E_{L_1}[M_{n_t}^{-1}], L_1 \gg L_2\), and only a small number, \(L_2\), of expensive inversions to calculate the small correction.

To check this method we compared the exact result for \((M^{-1})^{s1c1s2c2}_{xy}\), where \(s1c1\) denotes the spin and color indices, \(x = (0, 0, 0, 3)\) and \(y = (i, 0, 0, 3), i = 0 \ldots 10\), with an estimate obtained from eqn. \[^2\]. As expected we find consistency within errors for different \(n_t, L_1\) and \(L_2\). For example, for \(n_t = 5, L_1 = 5500, L_2 = 300, s1 = s2 = 1, c1 = c2 = 2\), \(E[M_{n_t}^{-1}] = (0.0300(7), -0.0014(7))\) compared to the exact result of \((0.0302 \ldots, -0.0010 \ldots)\).

We now have two parameters, \(n_t\) and the ratio \(L_1/L_2\), which need to be fixed, ideally, so as to minimize the variance of the disconnected loop, \(\text{Tr}(\Gamma M_{n_c}^{-1})\), at fixed cost. For \(L_1/L_2 \gg 1\) the variance (Var) is given by

\[
\text{Var}_{L_1} [\text{Tr}(\Gamma M_{n_t}^{-1})] + \text{Var}_{L_2} [\text{Tr}(\Gamma (M_{n_t}^{-1} - M_{n_c}^{-1}))] = \frac{f_1}{L_1} + \frac{f_2}{L_2},
\]

(1.4)

where \(f_1\) and \(f_2\) depend on \(n_t\) and \(\Gamma\), while the approximate cost is given by

\[
C = L_1 n_t + L_2 n_c.
\]

(1.5)

Using Lagrange multipliers and assuming \(f_1\) to be approximately independent of \(n_t\) we obtain the optimal values

\[
n_t^{\text{opt}} = \frac{1}{n_c} \frac{f_2 f_1}{(f_2 f_1)^2}, \quad \frac{L_1}{L_2} = \sqrt{\frac{f_1}{f_2} \frac{n_c}{n_t^{\text{opt}}}},
\]

(1.6)

where \(f_i^* = \partial f_i / \partial n_t\). For our observables we find that using these optimal values leads to \(L_1 \approx f_1^* / f_2^*\).

Additional gain can be obtained by combining with other noise reduction techniques. Here we consider the HPE approach \[^2, 5\]. The expansion of \(E[\text{Tr}(\Gamma M_{n_c}^{-1})]\) to order \(m\) is given by:

\[
E[\text{Tr}(\Gamma M_{n_c}^{-1})] = \frac{1}{L} \sum \left[ \langle \eta^l | \Gamma | \eta^l \rangle + \langle \eta^l | \Gamma \kappa D | \eta^l \rangle + \ldots + \langle \eta^l | \Gamma \kappa^m D^m | \eta^l \rangle \right] + E[\text{Tr}(\Gamma \kappa^{m+1} D^{m+1} M_{n_c}^{-1})],
\]

(1.7)
Disconnected contributions to hadronic structure

Sara Collins

where, since this is a geometric series, the last term gives the remainder, $\sum_{p=m+1}^{\infty} \langle \eta' | \Gamma \kappa^p \bar{D}^p | \eta' \rangle$, averaged over stochastic sources. One can omit terms in the expansion which only contribute to the noise. All odd terms, $\text{Tr}(\Gamma D^{2m+1}) = 0$, $\forall \Gamma$. For the even terms, $\text{Tr}(\Gamma) = 0$ $\forall \Gamma \neq 1$, while $1 \text{Tr}(\Gamma D^2) = 0$ $\forall \Gamma$ and for $\Gamma = \gamma_\mu \gamma_5$ and $\gamma_5$, even $\text{Tr}(\Gamma D^4) = \text{Tr}(\Gamma D^6) = 0$. Hence, for $\Gamma = \gamma_\mu \gamma_5$ and $\gamma_5$, since all terms up to 8th order only contribute to the noise, an improved estimate of the trace is given by $E[\text{Tr}(\Gamma \kappa^8 \bar{D}^8 \Gamma^{-1})]$. For all other $\gamma$ combinations we use $^2 E[\text{Tr}(\Gamma \kappa^4 \bar{D}^4 \Gamma^{-1})]$. The 4th and 6th order terms can be calculated explicitly $^2$ to achieve the same level of improvement as for $\Gamma = \gamma_\mu \gamma_5$ and $\gamma_5$, however, we have not done so in this study. To combine with our truncated solver method we substitute, for example, $\kappa^8 \bar{D}^8 \Gamma^{-1}$ for $\Gamma^{-1}$ in eqn. [3].

2. Results

We have performed an exploratory study of our method using configurations provided by the Wuppertal group: these are $n_f = 2 + 1$ dynamical configurations generated using a Symanzik improved gauge action and a stout-link improved staggered fermion action. The lattice spacing is fairly coarse, $a^{-1} \approx 1.55$ GeV while the volume is around 2 fm. Further details can be found in [9]. For valence quarks we used the Wilson action with $\kappa = 0.166, 0.1675$ and 0.1684 corresponding to pseudoscalar masses of about 600, 450 and 300 MeV respectively. Our main results were obtained using the conjugate gradient algorithm with even-odd preconditioning to perform the propagator inversions. However, section 2.3 will show results obtained using the stabilised biconjugate gradient algorithm (BiCGStab). The code used throughout was a modified version of the Chroma code [7].

Results are presented below for the disconnected loop, $\text{Tr}(\Gamma \Gamma^{-1})$, where we have considered $\Gamma = \Gamma$, $\gamma_\mu$, $\gamma_\mu \gamma_5$, $\sigma_{\mu \nu}$, $\gamma_5$. Using $\Gamma^{-1} = \gamma_5 (\Gamma^{-1})^\dagger \gamma_5$ one can show that the trace is either real or imaginary $^3$. At this initial stage we are only interested in the stochastic error and, hence, the results are presented for the trace on a single configuration. In addition to combining our method with the HPE approach we also partition in time: $|\eta'\rangle$ are only non-zero for $t = 3$.

2.1 Truncating the solver

The truncated solver method (TSM) relies on $\text{Tr}(\Gamma \Gamma_{n_i}^{-1})$ coming close to the convergent value after only a few iterations of the inverter. We found this to be true for all $\Gamma$ studied and the example of $\Gamma = \Gamma$ is shown in figure [4]. Clearly the trace is close to the limiting value after 20 iterations (compared to the 480 iterations needed for convergence). Proceeding to the calculation of the optimal values for $n_i$ and $L_1/L_2$, we use $\text{Tr}(\Gamma \Gamma_{n_i}^{-1})$ and $\text{Tr}[\Gamma (\Gamma_{n_i}^{-1} - \Gamma_{n_i}^{-2})]$ calculated on a single set of stochastic estimates, $L = 300$, for $n_i = 2$ to 100 in steps of 2 iterations to estimate $f_1$, $f_2$ and $f_2'$ as functions of $n_i$. Using eqn. [3] we obtain the optimal values given in table [4]. The results are presented for a subset of $\Gamma$s and show that all values for $n_i^{opt}$ are small, but also that $n_i^{opt}$ and $L_1/L_2$ depend on the $\Gamma$ used.

No error analysis has been attempted for these values and they should be considered rough estimates. However, we have increased the number of stochastic estimates to 500 and no significant

---

1These terms are zero for the Wilson action. For the clover action only the $m = 0$ term can be omitted.

2Where for $\Gamma = \Gamma$ we construct the non-vanishing $\text{Tr} \Gamma$.

3Of course the path integral expectation value $\langle \text{Tr} \Gamma^{-1} \rangle = 0$ $\forall \Gamma \neq \Gamma$. 

4
Disconnected contributions to hadronic structure
Sara Collins

Figure 1: The disconnected loop for $\Gamma = 1$ as a function of the number of iterations used in the inverter for $M^{-1}$ for $\kappa = 0.166$. The loop is shown for (left) $L = 1$ where the horizontal line shows the value at convergence ($n_c = 480$) and (right), with errors, $L = 300$.

Table 1: Optimal values for $n_t$ and $L_1/L_2$ for a subset of the $\Gamma$s studied, calculated using $L = 300$ for $\kappa = 0.166$. The gains obtained for the estimate of $\text{Tr}(\Gamma M^{-1})$ using these optimal values at fixed cost are also shown. Where our method is combined with the HPE technique, $m$ indicates the order used.

| $\Gamma$ | $1$ | $\gamma_5$ | $\gamma_1 \gamma_2$ | $\gamma_3 | \gamma_5$ | $1$ | $\gamma_3$ | $\gamma_1 | \gamma_2$ | $\gamma_3 | \gamma_5$ |
|----------|-----|-------------|---------------------|-----------------|-----|-------------|---------------------|-----------------|
| $n_t^{opt}$ | 50  | 27          | 14                  | 18              | 18  | 66          | 78                  | 50              | 78              | 90            |
| $L_1/L_2$ | 23  | 21          | 32                  | 28              | 30  | 26          | 25                  | 21              | 26              | 26            |
| $m$ | 4   | 4           | 4                   | 8               | 8   | 4           | 4                   | 8               | 8               | 8             |
| Gain | 5   | 5           | 10                  | 8               | 8   | 8           | 11                  | 19              | 24              | 29            |

change in the results was found. Using $n_t^{opt}$ and $L_1/L_2$ we can calculate the gain in computer time using the TSM at fixed cost this time with $L_1$ and $L_2$ independent stochastic sources. The cost, to be inserted in eqn. (1.5) is set by generating 300 stochastic estimates of $\text{Tr}(\Gamma M^{-1})$. The gain corresponds to

$$\text{Gain} = \frac{\text{Var}[\text{Tr}(\Gamma M^{-1})]}{\text{Var}[\text{Tr}(\Gamma M^{-1})]_{\text{[TSM]}}}. \quad \text{(2.1)}$$

Table 1 shows the TSM to result in significant gains for all $\Gamma$s studied, including $\Gamma = 1$. Note that, if time partitioning is not used in nominator and denominator these numbers are likely to be much larger.

We expect further variance reductions to be achieved when combining our method with the HPE technique discussed in section 1.2. Figure 2 shows the disconnected loop for $\kappa = 0.166$, which corresponds to about 20% below the strange quark mass, for $\Gamma = 1$ and $\gamma_3 \gamma_5$. We see that for $\Gamma = 1$ the variance does not reduce significantly as $\kappa D$ is applied up to the limit of $m = 4$. This is also the case for $\Gamma = \gamma_3$. However, for all other $\Gamma$s significant reductions in the variance are found, as seen for $\gamma_3 \gamma_5$.

Once combined with the TSM, the optimal values, $n_t^{opt}$ and $L_1/L_2$ must be recalculated. Table 1 shows that $n_t^{opt}$ increases compared to using TSM alone, however, it is still much less than...
Disconnected contributions to hadronic structure
Sara Collins

2.2 Effect of decreasing the quark mass

The results presented so far have been for a quark mass slightly below the strange quark mass. If the quark mass is reduced further, table 2 shows that down to \( m_{PS} \approx 300 \text{ MeV} \) there is no significant change in the values for the TSM method. As expected, the HPE technique becomes less effective as the quark mass decreases and this is reflected in the drop in the factors for the combined TSM+HPE. Nevertheless, at 300 MeV the gain is still \( \geq 2 \) times that for the TSM method alone for some of the \( \Gamma \)s.

2.3 Using a different solver

The results in the previous sections were obtained using the conjugate gradient (CG) algorithm in the solver. We are repeating the study using BiCGStab to see whether we can also achieve high gains with a more optimized solver. BiCGStab converges in less iterations than CG, for example, \( n_c = 156 \) compared to 480 for CG at \( \kappa = 0.166 \). However, each iteration is more expensive. Furthermore, BiCGStab does not converge smoothly. This means we cannot calculate optimal values for \( n_t \) and \( L_1/L_2 \) (which depend on \( \partial f_2/\partial n_t \)). However, we can fix \( L_1/L_2 \approx f_1/f_2 \) by

\[
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\text{Gain} & \text{TSM} & \text{TSM+HPE} \\
\hline
m_{PS} & \parallel & \gamma_3 & \gamma_2 & \gamma_5 & \gamma_3 & \gamma_1 & \gamma_2 & \gamma_5 & \gamma_3 & \gamma_1 & \gamma_2 & \gamma_5 \\
\hline
600 \text{ MeV} & 5 & 5 & 10 & 8 & 8 & 8 & 11 & 19 & 24 & 29 \\
450 \text{ MeV} & 5 & 5 & 10 & 8 & 8 & 7 & 11 & 17 & 22 & 24 \\
300 \text{ MeV} & 5 & 5 & 10 & 8 & 8 & 6 & 9 & 14 & 17 & 18 \\
\hline
\end{array}
\]
Disconnected contributions to hadronic structure

Sara Collins

requiring $\text{Var}[\text{Tr}(\Gamma M^{-1} n_t)] \approx \text{Var}[\text{Tr}(\Gamma (M^{-1} - M^{-1}_n))]$ and vary $n_t$ to find the best gain. Initial results using $n_t = 14$ give, for example, gains of 9 and 24 for $\Gamma = \gamma_3 \gamma_5$ using the TSM and TSM+HPE respectively, similar to the factors obtained using the CG solver.

3. Summary

The truncated solver method works well, providing gains in computer time of factors of $4 - 12$ for the disconnected loop, depending on the operator, for quark masses in the range of $m_{ps} = 600 - 300$ MeV. The method is easy to implement, independent of the quark action and, as we have shown, can be combined with other methods like the HPE technique to obtain gains of factors of around 30 for some operators. Future work will include combining our method with the truncated eigenmode approach and a study of the size of the gauge noise.

Acknowledgments

S. Collins gratefully acknowledges funding from the Stifterverband für die Deutsche Wissenschaft. This work has also been supported by the EC Hadron Physics I3 Contract RI3-CT-2004-506087, the BMBF Project 06RY258 and the DFG. We thank Mike Clark, Chris Michael and Hartmut Neff for discussions.

References


